

### **REMARKS**

The Office Action dated February 22, 2008 has been received and carefully noted. The above amendments to the claims, and the following remarks, are submitted as a full and complete response thereto.

Claims 1-4, 13 and 16 are amended to more particularly point out and distinctly claim the subject matter of the present invention. No new matter is added. Claims 1-21 are respectfully submitted for consideration.

The Office Action indicated that claims 16-18 contain allowable subject matter and would be allowed if rewritten in independent form. Applicants wish to thank the Examiner for the indication of allowable subject matter. However, claims 1-17 and 19-21 are respectfully submitted for reconsideration.

The Office Action rejected claims 1-3, 5-7, 9-11, and 13-15 under 35 U.S.C. 103(a) over US Patent No. 6,526,395 to Morris (Morris), in view of "Speech Recognition in Adverse Environments using Lip Information" to Thambiratnam et al (Tham). The Office Action alleged that Morris discloses all of the subject matter of the claims except processing the video signals based on a detection that at least a portion of the audio signal cannot be processed. The Office Action then alleged that Tham teaches this deficiency of Morris and that it would have been "obvious to try" processing the video signals based on a detection that at least a portion of the audio signal cannot be processed. Applicant submits that the Morris and Tham, taken individually or in combination, fail to disclose or suggest all of the features recited in any of the pending claims. Furthermore,

Applicant submits that it would not have been obvious to try processing the video signals based on a detection that at least a portion of the audio signal cannot be processed.

Claim 1, from which claims 2-4, 16, and 19 depend, is directed to a method of speech recognition. Audio signals are received from a speech source. Video signals are received from the speech source. It is determined if the audio signals can be processed. Based on the detection that at least a portion of audio signals can not be processed, the video signals are processed. At least one of the audio signals and the video signals are converted into recognizable information. A task is implemented based on the recognizable information.

Claim 5, from which claims 6-8, 17, and 20 depend, is directed to a speech recognition device. An audio signal receiver is configured to receive audio signals from a speech source. A video signal receiver is configured to receive video signals from the speech source. A processing unit is configured to detect if the audio signals can be processed and if so, to process the audio signals. The video signals are processed based on the detection that at least a portion of the audio signals cannot be processed. A conversion unit is configured to convert at least one of the audio signals and the video signals to recognizable information. An implementation unit is configured to implement a task based on the recognizable information.

Claim 9, from which claims 10-12, 18, and 21 depend, is directed to a system for speech recognition. A first receiving means is configured for receiving audio signals from a speech source. A second receiving means is configured for receiving video signals from

the speech source. A processing means is configured for detecting if the audio signals can be processed and processing the audio signals if the audio signals can be processed. The processing means processes the video signals based on the detection that at least a portion of the audio signals can not be processed. A converting means is configured for converting at least one of the audio signals and the video signals to recognizable information. An implementing means is configured for implementing a task based on the recognizable information.

Claim 13 is directed to a method of speech recognition. Audio signals are received from a speech source. Video signals are received from the speech source. If the audio signals can be converted into a recognizable format, the audio signals are processed. The audio signals are converted into recognizable information. The video signals are processed when a segment of the audio signals can not be converted into the recognizable information. The video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information. The processed video signals are converted into the recognizable information. A task is implemented based on the recognizable information.

Claim 14 is directed to a speech recognition device. An audio signal receiver is configured to receive audio signals from a speech source. A video signal receiver is configured to receive video signals from the speech source. A first processing unit is configured to detect if the audio signals can be converted, and if the audio signals can be converted, the audio signals are processed. A first conversion unit is configured to

convert the audio signals to recognizable information. A second processing unit is configured to process the video signals when the audio signals cannot be converted into the recognizable information. The video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information. A second conversion unit is configured to convert the processed video signals into the recognizable information. An implementation unit is configured to implement a task based on the recognizable information.

Claim 15 is directed to a system for speech recognition. A first receiving means receives audio signals from a speech source. A second receiving means receives video signals from the speech source. A first processing means detects if the audio signals can be converted, and if the audio signals can be converted, the audio signals are processed. A first converting means converts the audio signals into recognizable information. A second processing means processes the video signals when a segment of the audio signals can not be converted into the recognizable information. The video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information. A second converting means converts the processed video signals into the recognizable information. An implementing means implements a task based on the recognizable information.

Applicants submit that each of the pending claims recites features that are neither disclosed nor suggested in the Morris and Tham.

Morris is directed to an apparatus which includes a video input unit and an audio input unit. The apparatus also includes a multi-sensor recognition unit coupled to the video input unit and the audio input unit, and a processor coupled to the multi-sensor recognition unit. The multi-sensor recognition unit decodes a combined video and audio stream containing a set of user inputs. Fig. 2 of Morris illustrates input interpretation of the system. According to Morris the audio input and video input are processed in parallel. See Fig. 2, refs. 206, 208, 210 and 212. In block 210, visual gestures are captured from the speech input unit (See col. 4, lines 25-44).

The Office Action admits that Morris failed to disclose the feature of “processing the video signals if it is detected that at least a portion of the audio signal cannot be processed”, as recited, in part, in independent claim 1, and similarly in independent claims 5, 9 and 13-15. The Office Action relied on Tham and what is allegedly “obvious to try” to cure the deficiencies of Morris with respect to independent claims 1, 5, 9 and 13-15.

Tham does not cure the deficiencies of Morris with respect to independent claims 1, 5, 9 and 13-15. Tham discloses a speech recognition system that integrates audio voice signals and video signals taken of human lips. The system uses a lip tracking mechanism called abstract shape models (ASMs) to perform lip tracking and parameterization which uses a hidden Markov model (HMMs) to perform the recognition.

Integration of the video and audio is performed by one of two primary methods of integration which are direct integration and asynchronous integration. For direct

integration, the audio and video vectors are combined as input to a recognizer (see Fig. 2 of Tham). For asynchronous integration, the data is merged as system output based on the two results which are calculated independent of one another. The results of the two systems are converted to probabilities which are combined into a single probability. One benefit of the asynchronous integration model is that the results can be independently determined and can use different frame rates.

In Tham, both the direct integration and asynchronous integration systems for speech recognition fail to disclose “processing the video signals if it is detected that at least a portion of the audio signal cannot be processed”, as recited, in part, in independent claim 1, and similarly in independent claims 5, 9 and 13-15. The Office Action alleged that,

“Fundamentally, one having ordinary skill in the art would readily understand that a speech recognizer that utilizes both audio and video for purposes of recognition would utilize the video if the quality of the audio information is poor, and utilize the audio if the quality of the audio information is good.”

Applicant disagrees, and submits that a *prima facie* case of obviousness has not been made because all of the claim limitations have not been taught by the combination of references cited, namely, Morris and Tham. 35 U.S.C. §103(a) requires that all of the claim limitations be taught by a combination of references in order to establish a *prima facie* case of obviousness. The Office Action has failed to teach all of the claim limitations as the Office Action has failed to provide a reference that teaches “processing

the video signals if **it is detected that** at least a portion of **the audio signal cannot be processed**", as recited, in part, in independent claim 1, and similarly in independent claims 5, 9 and 13-15.

The Office Action further alleged that the combined probability of the asynchronous integration method disclosed in Tham supports that the "processing the video signals if it is detected that at least a portion of the audio signal cannot be processed", as recited, in part, in independent claim 1, and similarly in independent claims 5, 9 and 13-15. Applicant disagrees and submits that the mere combining of audio and video results does not substitute processing video signals **if** it is determined that an **audio signal cannot be processed**, as recited by claim 1.

The Office Action then alleged that it would have been "obvious to try" "processing the video signals if it is detected that at least a portion of the audio signal cannot be processed", as recited, in part, in independent claim 1, and similarly in independent claims 5, 9 and 13-15. Applicant disagrees and submits that the case law standard of what is "obvious to try", as supported by §2143 of the MPEP (see page 2100-134 of the MPEP, Rev. 6, Sept. 2007) was misapplied by the Office Action and is not relevant to the subject matter recited in the claims.

The standard for what is obvious to try is described in detail in the MPEP. §2143-(E) of the MPEP explicitly requires that in order for a claim limitation to be "obvious to try" the Office personnel must resolve the *Graham* factual inquiries by articulating certain requirements.

The first (1) requirement is a finding that at the time of the invention, there had been a recognized problem or need in the art, which may include a design need or market pressure to solve a problem. Applicant submits that there is no prior established “recognized problem” of substituting video for audio if the audio cannot be processed, as recited in claim 1. Further, there is no prior established “design need” or “market pressure” for substituting video for audio if the audio cannot be processed, as recited in claim 1.

The second (2) requirement is a finding that there had been a finite number of identified, predictable potential solutions to the recognized need or problem. There is no finite number of identified or predictable solutions related to the design decision for substituting video for audio if the audio cannot be processed, as recited in claim 1.

Furthermore, the MPEP provides three (3) examples of what would have been “obvious to try” and which fit the requirements of the above noted requirements. Applicant submits that none of these three known solutions are applicable to the subject matter of the claimed invention.

The first (1<sup>st</sup>) relates to a pharmaceutical drug that had a known result that was predictable based on a suitable salt to be used in a drug to be administered to a human. The list of different salts that could be used was finite and the particular salt selected was one of fifty-three different possibilities, thus representing a finite number of possible solutions to a problem.



The second (2<sup>nd</sup>) example relates to a formula for a sustained-release pharmaceutical drug that had a known release period of twenty-four (24) hours and was created based on a known time interval being the solution sought based on a known chemical Oxybuynin with known chemical properties such as high water solubility.

The third (3<sup>rd</sup>) example discloses a procedure used to isolate a nucleic acid molecule. There are only a few known way to isolate a nucleic acid molecule, and thus it is obvious to try those few known methods to achieve a specific result.

Therefore, Applicant submits that for at least the reasons stated above, Morris, Tham and the rationale for what is “obvious to try” fail to teach “processing the video signals if it is detected that at least a portion of the audio signal cannot be processed”, as recited, in part, in independent claim 1, and similarly in independent claims 5, 9 and 13-15. By virtue of dependency claims 2-4, 6-8, 10-12 and 16-21 are also allowable.

Accordingly, all of the claim limitations of independent claims 1, 5, 9 and 13-15 have not been taught or suggested by the prior art, and the rejection of claims 1-3, 5-7, 9-11 and 13-15 is improper and must be withdrawn.

The Office Action rejected claims 4, 8 and 12 under 35 U.S.C. 103(a) as being obvious over Morris and Tham, in view of US Patent No. 6,219,639 to Bakis et al. (Bakis). The Office Action took the position that Morris and Tham disclosed all of the features recited in these claims except storing audio and video signals to a destination source and a transmitter for sending the audio signals and the video signals to a destination source. The Office Action asserted that it is well-known in the art to operate

biometric identification via a client/server network, where biometric data is stored on a server and biometric data is collected locally and compared to stored biometric data on the server. The Office Action relied on Bakis in support of this assertion. Applicants submit that the cited references, taken individually or in combination, fail to disclose or suggest all of the features recited in any of the pending claims. Specifically, Morris and Tham are deficient at least for the reasons discussed above, and Bakis fails to cure these deficiencies.

Bakis is directed to recognizing an individual based on attributes associated with the individual. Bakis describes pre-storing previously extracted biometric attributes which may be later retrieved for the purpose of comparison with subsequently extracted biometric attributes to see if a match exists between the two. See col. 8 lines 47-53.

However, Applicants submit that Bakis is merely a cumulative reference when combined with Morris. The combination discloses comparing biometric data with stored biometric data. Thus, Bakis fails to cure the significant deficiencies of Morris and Tham discussed above.

Based at least on the above, Applicants respectfully submit that the cited references fail to disclose or suggest all of the features recited in claims 4, 8 and 12. Accordingly, withdrawal of the rejection under 35 U.S.C. 103(a) is respectfully requested.

The Office Action rejected claims 19-21 under 35 U.S.C. 103(a) as being obvious over Morris and Verma, in further view of US Patent No. 5,412,738 to Brunelli et al.

(Brunelli). The Office Action took the position that Morris and Tham disclosed all of the features of these claims except determining if the images of the user are detected and indicating to the user if the video image is not detected. The Office Action asserted that Brunelli disclosed these features. Applicants respectfully submit that the cited references, taken individually or in combination, fail to disclose or suggest all of the features recited in any of the above claims. Specifically, Applicants submit that Morris and Tham are deficient at least for the reasons discussed above, and Brunelli fails to cure these deficiencies.

Brunelli is directed to a recognition i.e., identification and verification, system. Acoustic and visual features are integrated to identify people and to verify their identities. However, Applicants respectfully submit that Brunelli fails to cure the significant deficiencies of Morris and Tham discussed above regarding claims 1, 5, and 9.

Further, Applicants respectfully submit that Brunelli fails to cure the admitted deficiencies of Morris and Tham. As discussed above, the Office Action relied on Brunelli to disclose the feature of determining if the images of the user are detected and indicating to the user if the video image is not detected. The Office Action further alleged that a person is notified when his/her image are not detected because an acoustic indicator does not prompt the user the speak the words, so the absence of a prompt is equivalent to an indication that the video image was not detected. Applicants respectfully submit that the Office Action is inappropriately reading features into Brunelli. First, the features recited in claims 19-21 are positive steps thus an “indication” is not given that

the video image is not detected. Further, the Office Action has not provided evidence that the absence of a prompt is necessarily a positive indication that the video image is not detected because the absence of a prompt could be based on any number of factors, does not necessarily flow from the non-detection of the video image. Thus, Brunelli fails to cure the admitted deficiencies of Morris and Tham.

Based at least on the above, Applicants respectfully submit that the cited references fail to disclose or suggest all of the features recited in claims 19-21. Accordingly, withdrawal of the rejection under 35 U.S.C. 103(a) is respectfully requested.

Applicants submit that each of claims 1-21 recites features that are neither disclosed nor suggested in any of the cited references. Accordingly, it is respectfully requested that each of claims 1-21 be allowed, and this application passed to issue.

If for any reason the Examiner determines that the application is not now in condition for allowance, it is respectfully requested that the Examiner contact, by telephone, the applicant's undersigned attorney at the indicated telephone number to arrange for an interview to expedite the disposition of this application.

In the event this paper is not being timely filed, the applicant respectfully petitions for an appropriate extension of time. Any fees for such an extension together with any additional fees may be charged to Counsel's Deposit Account 50-2222.

Respectfully submitted,



---

Kamran Emdadi  
Registration No. 58,823

**Customer No. 32294**  
SQUIRE, SANDERS & DEMPSEY LLP  
14<sup>TH</sup> Floor  
8000 Towers Crescent Drive  
Tysons Corner, Virginia 22182-2700  
Telephone: 703-720-7800  
Fax: 703-720-7802

KE/jeh:cqc